Evaluation of Complex Distributed Multimodal Applications: Evaluating a TeleRehabilitation System When It Really Matters

Carlos Pereira[1,2(✉)], Nuno Almeida[1,2], Ana Isabel Martins[1,2], Samuel Silva[1], Ana Filipa Rosa[1], Miguel Oliveira e Silva[1,2], and António Teixeira[1,2]

[1] Institute of Electronics and Telematics Engineering, University of Aveiro, Aveiro, Portugal
cepereira@ua.pt

[2] Department of Electronics, Telecommunications and Informatics Engineering, University of Aveiro, Aveiro, Portugal

Abstract. The evaluation of applications or systems within dynamic environments is complex. The existence of multiple hardware and software items which share the same space can provoke concurrency issues and result in erratic interactions. A sudden change within the environment can result is dramatic changes both to the user and application itself which can pass unnoticed in traditional evaluation methodologies. To verify if a component is compatible with a given environment is of paramount importance for areas like pervasive computing, ambient intelligence or ambient assisted living (AAL). In this paper, a semi-automatic platform for evaluation is presented and integrated with a TeleRehabilitation system in an AAL scenario to enhance evaluation. Preliminary results show the advantages of the platform in comparison with typical observation solutions mainly in terms of achieved data and overall ease of use.

Keywords: Evaluation · Multimodality · TeleRehabilitation

1 Introduction

The design and development of complex multimodal systems, working in multiple devices and deployed in dynamic environments, poses several challenges. Beyond the technical aspects, designing user experience in this context is far from being simple. At this level, tasks and interaction modalities cannot be looked at as isolated phenomena [6]. For example, the use of several modalities simultaneously, as a result of a more complex use of the system, might result in sensory overload [9]; or particular modalities, which in abstract seem suitable options, are disregarded by the user in some (e.g., stressful) situations. Furthermore, these concerns are particularly relevant when the target users might present some level of disability, physical or cognitive, which directly influences how they use the system: an audio warning might not be heard by the user, due to a hearing disability, or multiple tasks crossing might leave the user disoriented [3]. Therefore, integration of proper evaluation, in the development cycles,

J. Zhou and G. Salvendy (Eds.): ITAP 2015, Part II, LNCS 9194, pp. 146–157, 2015.
DOI: 10.1007/978-3-319-20913-5_14

covering different contexts of use and complex tasks, running in its intended (real or simulated) environment, is of paramount importance and should be increasingly introduced, from early on, as a tool to support the development of such systems.

In this article, we present a semi-automatic evaluation platform and its usage applied to the assessment of a TeleRehabilitation system. This platform allows the creation of dynamic evaluation plans and, by continuously assessing context and user performance, provides evaluators with a more complete report of the experience. The collected data also allows inferring the precise timings to trigger questions to assess user performance and/or satisfaction, for example right after failing to complete a task or when the user is idle. The platform distinguishes itself from alternative tools [4,8] by using an ontology at its core and by providing a decoupled manner in which evaluators can integrate other software with it. Additionally, a dedicated user interface (UI) simplifies the creation and deployment of evaluation tests without requiring specific programming knowledge.

2 Supporting the Evaluation of Multimodal Distributed Systems - the DynEaaS Platform

Usual evaluation methods do not fully serve the task of assessing user feedback and performance in regard to applications in highly dynamical environments. Ubiquitous or pervasive systems require adaptable evaluation solutions that do not limit the amount of gathered data. By gathering additional data, evaluators are able to assess a wide range of aspects regarding their applications and take into account the surrounding environment when establishing conclusions.

Dynamic Evaluation as a Service (DynEaaS) (Fig. 1) is an evaluation platform capable of evaluating user performances in dynamical environments by allowing evaluation teams to create and conduct context-aware evaluations. The platform allows evaluators to specify evaluation plans which contain actions that are triggered at precise timings or only when certain conditions are met, thus gathering better contextualized data.

DynEaaS follows a distributed paradigm allowing the evaluator to run multiple evaluations at different locations simultaneously. At each location, the plan is instantiated and applied taking into account user preferences, current context and the environment itself. When applying the plan, DynEaaS constantly evaluates the current context and chooses the best suited conditions to interact with the user.

Within DynEaaS, each user is seen as a user node named EaaS Node which is a part of an evaluation network. An evaluation can be remotely started for a defined set of users. Each evaluation network is defined by a set of criteria which every user node must comply with and is controlled by a central node called EaaS Core. Some examples of criteria can encompass user preferences or interests as well as more structural aspects such as hardware or environment conditions.

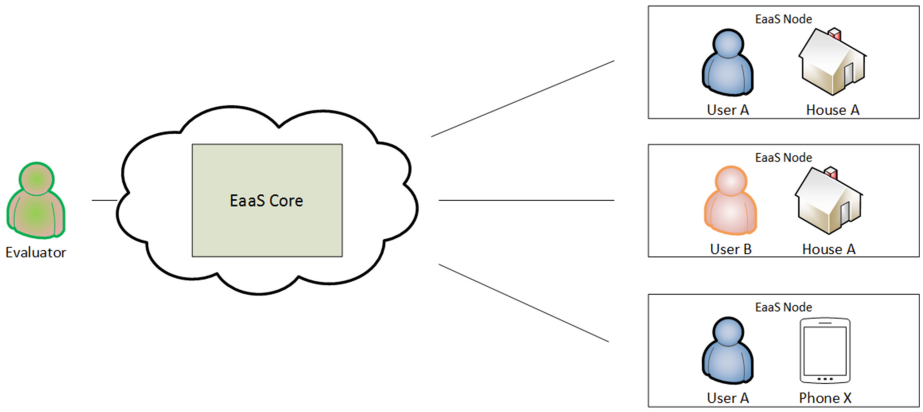


Fig. 1. DynEaaS ecosystem

Results are synchronized in real time. By having access to them, the evaluator is able to analyze current data and have a better grasp over the evaluation current status making small changes to it, if so required.

DynEaaS embraces a International Classification of Functioning, Disability and Health (ICF) [12] based methodology that includes different usability evaluation methods such as questionnaires and performance evaluation. The environmental factors are a central aspect of the ICF based methodology. Using the DynEaaS platform it is possible to assess every situation foreseen by the evaluator concerning a system, a user or the entire environment by defining events triggering actions (e.g., questions). Events can encompass temporal aspects (specific times), environmental aspects (noise, brightness), contextual aspects (persons in the room, interruptions), interaction options (repeated actions), among others, which can be aggregated to create specific evaluation contexts. All data is recorded and can be further analyzed later.

By using ontologies, the platform is highly flexible and can be used in different domains without core changes. Post-evaluation reasoning operations are also possible if required.

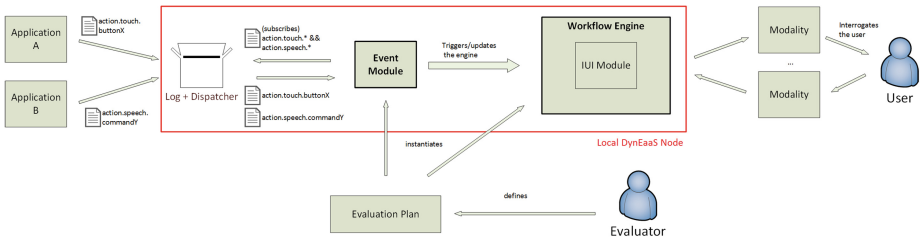


Fig. 2. DynEaaS local architecture exemplification

Locally, each EaaS Node (illustrated in Fig. 2) is composed of a set of services which cooperate to execute evaluation plans. These evaluation plans are created by the evaluator at EaaS Cores and deployed on-demand to selected user nodes.

Each evaluation plan encompasses an unbounded number of workflows with the objective of gathering specific information from the user. Each workflow can be seen as a tree which is started at its root and executed until it reaches all of its leafs. These workflows are executed by the Workflow Engine within the EaaS Node and delivered to the user via associated modalities/user interfaces. The selection of which modality to use falls on the IUI Module based on current context and the evaluation specifications.

Each workflow can contain two types of elements: event rules and inquiries. An inquiry comprises a number of questions to be asked in succession. DynEaaS supports both open-answer questions and multi-answer questions which are created by the evaluator. Event rules on the other hand enable the creation of complex event compositions.

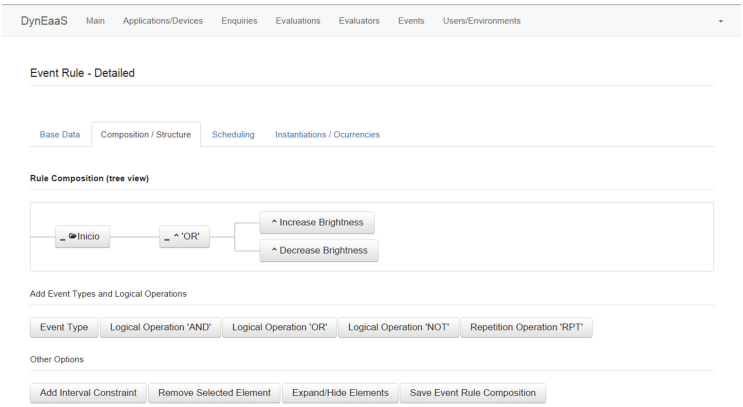


Fig. 3. DynEaaS UI for creating event rules dynamically

In DynEaaS, each event is described by an EventType which defines a routing key for it. These routing keys are used by applications to deliver notifications to DynEaaS using a decoupled message queue (Log+Dispatcher). This message queue is associated with an Event Module which receives selected events (according to active workflows which trigger the engine). EventTypes can also be used to form EventRules using a number of operators such as:

- 'And' and 'Or' Operator - creates a logical operation between two elements (either types or other operators)
- 'Not' Operator - negates an element
- Delay Operator - waits a period of time after or before evaluating an element
- Functor Operator (such as BiggerThan or SmallerThan) - enables the creation of predicate functions that compare arguments inside the events.

Each of these operators (except functor) can be applied to other operators which makes the creation of event rules limitless. The platform is accessible via a graphical UI which enables evaluators to create, design and deploy an evaluation plan to any linked user. Figure 3 demonstrates the UI for the creation of a simple event rule that triggers when either an increase or a decrease on brightness occurs.

3 Evaluation of TeleRehabilitation Using DynEaaS

3.1 TeleRehabilitation Application

TeleRehabilitation [11] is a new service which allows a patient to have a remote session of rehabilitation with a physiotherapist. The system provides different features for the patient and for the physiotherapist (Fig. 4).

Figure 5 illustrates the telerehabilitation system with both user interfaces. On the patient side, the application is divided in four major components: live video of the user doing the exercises, video presentation illustrating the current exercise, state of the session, e.g., duration, and a chat window. On the physiotherapist side, the application is divided in five components: exercise plan creation, plan status, vital signs monitor, live video of the patient and chat. TeleRehabilitation supports multimodal interaction [1,2], based on the W3C multimodal architecture [5], allowing the user to interact by touch and speech, as input modalities, and onscreen graphics and voice as output. Since the patient will be doing the exercises, and is far from the screen, speech interaction will be, most likely, the preferred modality.

With the application reaching an advanced stage of development, one of the challenges is how to perform the evaluation of TeleRehabilitation so that it

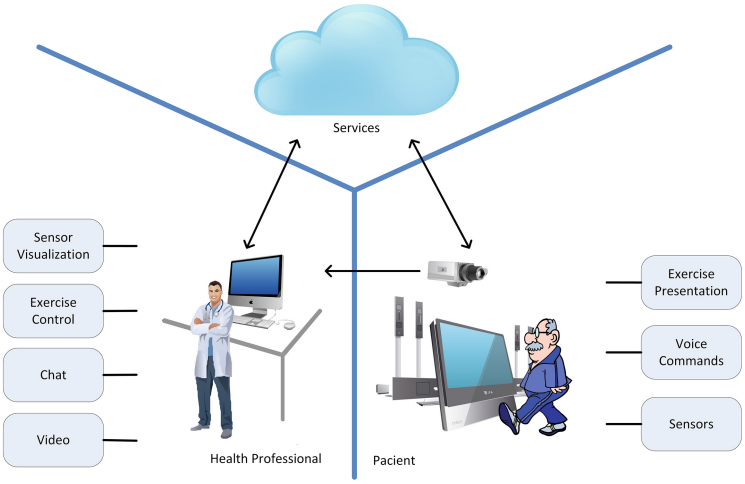
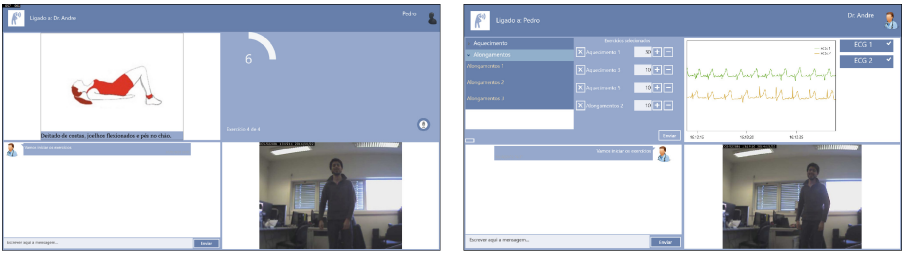


Fig. 4. TeleRehabilitation system architecture



(a) Patient Interface (b) Physiotherapist Interface

Fig. 5. The TeleRehabiliation application user interfaces

can encompass the full complexity of the system, its tasks and true multimodal interaction.

3.2 Specifying the Evaluation Protocol

The TeleRehabilitation system consists of two different modules addressing the two user profiles involved: therapists and patients. This fact makes it more difficult to evaluate the system as a whole, given that it is necessary to perform two evaluations at the same time, one for the patient and one for the physiotherapist.

The system itself requires validation from both users regarding its overall functioning and usability. To do so, we have used the DynEaaS platform with two user nodes, one for the patient and one for the physiotherapist.

In order not to obstruct the user, but at the same time obtain information in precise timings, we have embedded evaluation specific interfaces within both application modules. This way, the user is not required to shift his/her attention from the application and is able to insert information in real time. For this evaluation, users were able to interact using touch, keyboard and mouse or voice commands. The test was set within our Living Lab [10](in the case of the user) and a private room (in the case of the physiotherapist).

The evaluation session itself was composed of several exercises, activated by the therapist and displayed at the user's side. The user was asked to perform the exercises while the therapist observed and sent feedback. The performed sessions had an average duration of 15 min. Each test was accompanied by an evaluator to provide the necessary initial explanation to the user. The evaluator was also asked to compose a critical incidents registration.

Previously to the start of the evaluation, involved event types were specified in DynEaaS (see Table 1). In this case, only events produced by the TeleRehabilitation system have been inserted.

The list of possible events includes simple events like login or session start as well as more specific events such as sending a chat message, receiving current exercise status information or selecting a new exercise.

Based on these event types and also event rules and inquiries, we have created two plans, one for each intervenient. Both plans were set to start at the

Table 1. Event Listing

Event designation	Description
TR_NextExercise	Advancement to a new exercise.
TR_newExerciseList	Activation of a new exercise list.
TR_SensorSelect	Selection of a sensor in the interface.
TR_SendExercise	Sending a message to the other user.
TR_RemoveExercise	Elimination of an exercise from the exercise list.
TR_Login	Successfull authentication.
TR_PreviousExercise	Selection of a previous exercise.
TR_TimeChangeExercise	Changing the time period of an exercise.
TR_SensorZoomReset	Reseting the sensor viewing component.
TR_ExerciseStatus	Indication of the status of the currently active exercise.
TR_SensorZoom	Performing zoom on the sensor component.
TR_ReceiveMsgChat	Receiving a message on the chat component.
TR_SendMsgChat	Sending a message via chat component.
TR_SelectExercise	Selecting an exercise.
TR_endExerciseList	Finishing an exercise list.
TR_SessionStart	Initiating an exercise session.

same time. In order to obtain information from the user with questions, the plan integrates a set of evaluation flows. An evaluation flow depicts a set of linked event rules and inquiries with a specific order. When the evaluation is initialized, evaluation flows are instantiated into workflows and executed in the EaaS Nodes. Following are some examples which illustrate the diversity of the inserted evaluation flows:

- a flow intended to assert the overall opinion of the application after a certain usage. The flow is composed of two elements, the first being an event rule which triggers ten minutes after login, and the second, a question composed of a number of possible answers.
- a flow to assert a possible malfunctioning with the chat component. In case the user presses the 'Send chat message' five times in a row under ten seconds, a question is triggered asking the user the cause of that event. Note that while the first flow will occur due to being associated with time, the probability of this second flow happening is very slim. This helps demonstrate the flexibility of DynEaaS in the sense that it can support both time specific flows like the first while flows like the second one can help depict faults within the application itself.
- a flow to trigger a question when the user surpasses thirty percent of the exercise list. The question itself interrogates the user regarding the exercise demonstrations and its utility.

- a flow to trigger if the user has not used the chat functionality at all after ten minutes. In this case, the user is asked why did he not use that functionality, either by not noticing, not needing it or feeling it is not important.
- another flow operates similarly in regard to voice commands after five minutes.

On the physiotherapist side, we have created another set of evaluation flows, some similar to the ones for the patient (like overall impression of the system), and others to assert specific components on the physiotherapist side (such as inquiring why the physiotherapist did not use certain features).

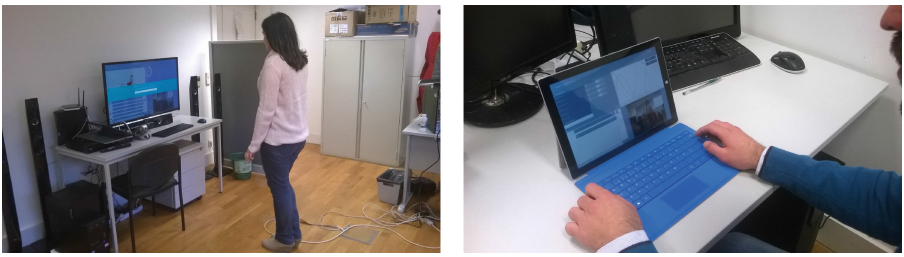
Overall, both plans aim at gathering information from a single therapy session, covering both sides of application use simultaneously (therapist and patient).

4 Results

Before performing a high number of evaluation sessions, a preliminary test was performed that, besides collecting evaluation data, should serve to validate the methodology, integration of DynEaaS with the application and assess the relevance of the defined evaluation plans. On this test, we have prepared a room for the therapist, and another for the patient and ran it for 15 min. An observer was present to take notes on the session from the patient side, and DynEaaS was placed running on both users.

Figure 6 shows both users during the initial test of the system. At the end of this test we reached two main conclusions. First, both DynEaaS plans were not extracting as much information as we desired. The truth was that the created plans were small. Initially, we feared that a high number of evaluation flows could constantly disrupt/distract the user from using the system and, therefore, created evaluation plans with a very limited scope. However, given the gathered results, we found that half of the actual workflows within both plans were dependent of the user himself and did not activate because the user had not fulfilled the necessary conditions.

The second conclusion was that the embedded interfaces within the system did not suit the system itself. Asking the user to answer the question with a keyboard and mouse lowered the user's usability dramatically.



(a) User View (b) Physiotherapist View

Fig. 6. A TeleRehabiliation session

Taking this in mind, we prepared a second trial which complemented these aspects. To tackle the issue with keyboard and mouse, we inserted a new interface into TeleRehabilition which enabled the user to answer all DynEaaS solicitations using speech.

4.1 Analyzing Obtained Results

In both sessions, the number of triggered workflows was very similar. Both users performed the login step, the requested exercises and used the available interaction modalities successfully. In session one, the number of detected events however was much smaller than in session two, especially concerning the chat component which indicated that patient two interacted with the therapist more often.

Either in session one and two, the therapist created a 12 step exercise list for the patient to complete which allowed us to compare timings between both. Results from DynEaaS showed that when reaching the 30 % percent of exercises performed, both patients found the exercise component to be clear and helpful in regard to its demonstration.

The results captured by the DynEaaS show good acceptance by both users. They were engaged within the rehabilitation session and indicated that they were satisfied with the interaction. These results are in compliance with previous usability evaluation test used to assess the preceding versions of the Telerehabilitation [7]. The previous test was made according with the traditional approach. The users followed a session script while the evaluator observed and collected data about the interaction. At the end of the session, the users completed a usability questionnaire regarding the system.

Evaluators claimed that when comparing the results of the evaluation made with DynEaaS and the previous one made with a traditional approach, it is possible to understand the practical value of DynEaaS as with a minor effort the evaluator has access to a greater amount of data regarding user experience.

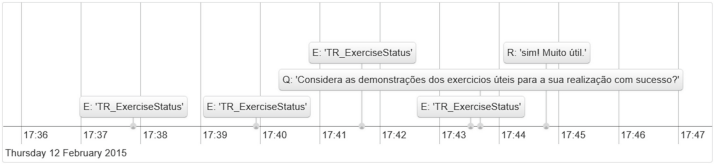


Fig. 7. DynEaaS UI showing results for a workflow in a timeline

An example of the added value to the evaluation process is the timeline generated from DynEaaS (Fig. 7). The figure illustrates a timeline generated from DynEaaS which allowed us to check the times that both patients took when performing the '30 % percent' workflow. Another interesting result that DynEaaS provided, concerns the overall opinion of the user after 10 min using

the system. Figure 8 presents two graphics, the first regarding the first session and the second concerning the second. The major difference from the first session to the second was the inclusion of speech. Results show that the amount of time that the patient took to provide the necessary user credentials (marked by 'Login' in Fig. 8) decreased by half from the first to the second session. This is also true when analyzing the timings for answering a question in the same workflow. Based on these results, we were able to confirm the importance and usefulness of speech.

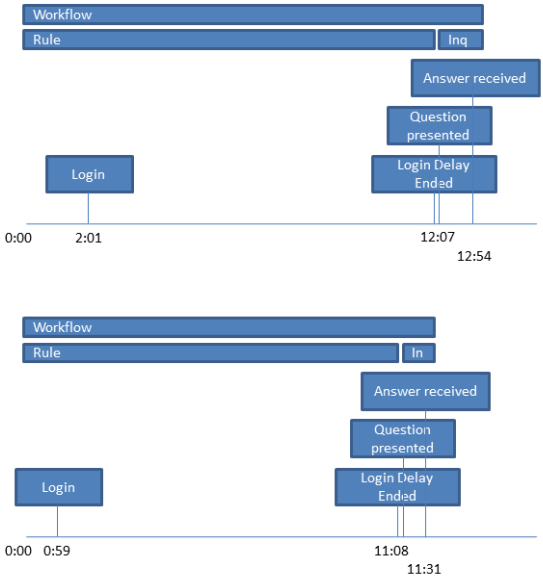


Fig. 8. Comparison between the two users concerning a specific workflow

In both sessions, observation reports indicated that the test itself went accordingly to what was expected, which also indicated that the user was engaged by the application.

5 Conclusions

The usage and flexibility of DynEaaS for evaluating the TeleRehabilitation system was proved to be very helpful. The creation of automatic workflows which trigger according to specific events allowed the extraction of valuable information from the user in real situations. For instance, DynEaaS allowed us to verify that the patient, when confronted with the ability to use speech interfaces, does so, even when speech recognition is not perfect.

In the future we intend to create more test cases with a higher number of users. While the presented case study was performed singularly and without concurrent applications, the compatibility of other software and hardware elements should be also asserted, preferably in non-controlled environments.

Additionally, the application of DynEaaS in other scenarios is also an objective as well as its enhancement, majorly by exploring the automatic generation of inquiries based on domain specific ontologies.

Acknowledgments. Research partially funded by IEETA Research Unit funding (PEst-OE/EEI/ UI0127/2014), project Cloud Thinking (funded by the QREN Mais Centro program, ref. CENTRO-07-ST24-FEDER-002031), Marie Curie Actions IRIS (ref. 610986, FP7-PEOPLE-2013-IAPP), and project AAL4ALL (ref. QREN 13852).

References

1. Almeida, N., Teixeira, A.: Enhanced interaction for the elderly supported by the w3c multimodal architecture. In: Conferência Nacional sobre Interacção. Vila Real, Portugal (2013)
2. Almeida, N., Silva, S., Teixeira, A.: Design and development of speech interaction: a methodology. In: Kurosu, M. (ed.) HCI 2014, Part II. LNCS, vol. 8511, pp. 370–381. Springer, Heidelberg (2014)
3. Blythe, M.A., Monk, A.F., Doughty, K.: Socially dependable design: The challenge of ageing populations for hci. Interact. Comput. **17(6)**, 672–689 (2005). http://dx.doi.org/10.1016/j.intcom.2005.09.005
4. Carter, S., Mankoff, J., Heer, J.: Momento: Support for situated ubicomp experimentation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 125–134. ACM, New York (2007). http://doi.acm.org/10.1145/1240624.1240644
5. Dahl, D.: The w3c multimodal architecture and interfaces standard. J. Multimodal User Interfaces **7**(3), 171–182 (2013)
6. Hollender, N., Hofmann, C., Deneke, M., Schmitz, B.: Review: Integrating cognitive load theory and concepts of human-computer interaction. Comput. Hum. Behav. **26(6)**, 1278–1288 (2010). http://dx.doi.org/10.1016/j.chb.2010.05.031
7. Martins, A.I., Queirós, A., Cerqueira, M., Alvarelhão, J., Teixeira, A., Rocha, N.: Assessment of Ambient Assisted Living Services in a Living Lab Approach: a Methodology based on ICF (2012)
8. Meschtscherjakov, A., Reitberger, W., Tscheligi, M.: Maestro: Orchestrating user behavior driven and context triggered experience sampling. In: Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research, MB 2010, pp. 29:1–29:4. ACM, New York (2010). http://doi.acm.org/10.1145/1931344.1931373
9. Oviatt, S., Coulston, R., Lunsford, R.: When do we interact multimodally?: Cognitive load and multimodal communication patterns. In: Proceedings of the 6th International Conference on Multimodal Interfaces, ICMI 2004, pp. 129–136. ACM, New York (2004). http://doi.acm.org/10.1145/1027933.1027957
10. Teixeira, A., Rocha, N., Pereira, C., et al.: The living usability lab architecture: Support for the development and evaluation of new aal services for the elderly. In: Ambient Assisted Living Book. CRC Press (2011)

11. Teixeira, A., Pereira, C., e Silva, M.O., Alvarelhao, J., Silva, A., Cerqueira, M., Martins, A.I., Pacheco, O., Almeida, N., Oliveira, C., Costa, R., Neves, A.J.R., Queiros, A., Rocha, N.: New telerehabilitation services for the elderly. In: Isabel Maria Miranda, M.M.C.C. (ed.) Handbook of Research on ICTs for Healthcare and Social Services: Developments and Applications. IGI Global (2013)
12. World Health Organization: International Classification of Functioning, Disability and Health (2001)